



University of
Nottingham

UK | CHINA | MALAYSIA



University of
BRISTOL

A large, high-resolution image of the Earth as seen from space, showing the Western Hemisphere. The Earth is centered in the frame, with the Atlantic Ocean, North America, and South America visible. The image is set against a dark, starry background of space. A white rectangular border is superimposed over the Earth image, containing the main text and logos.

Levelling-up via CIVIC and shopping data

N/LAB

Digital
Footprints

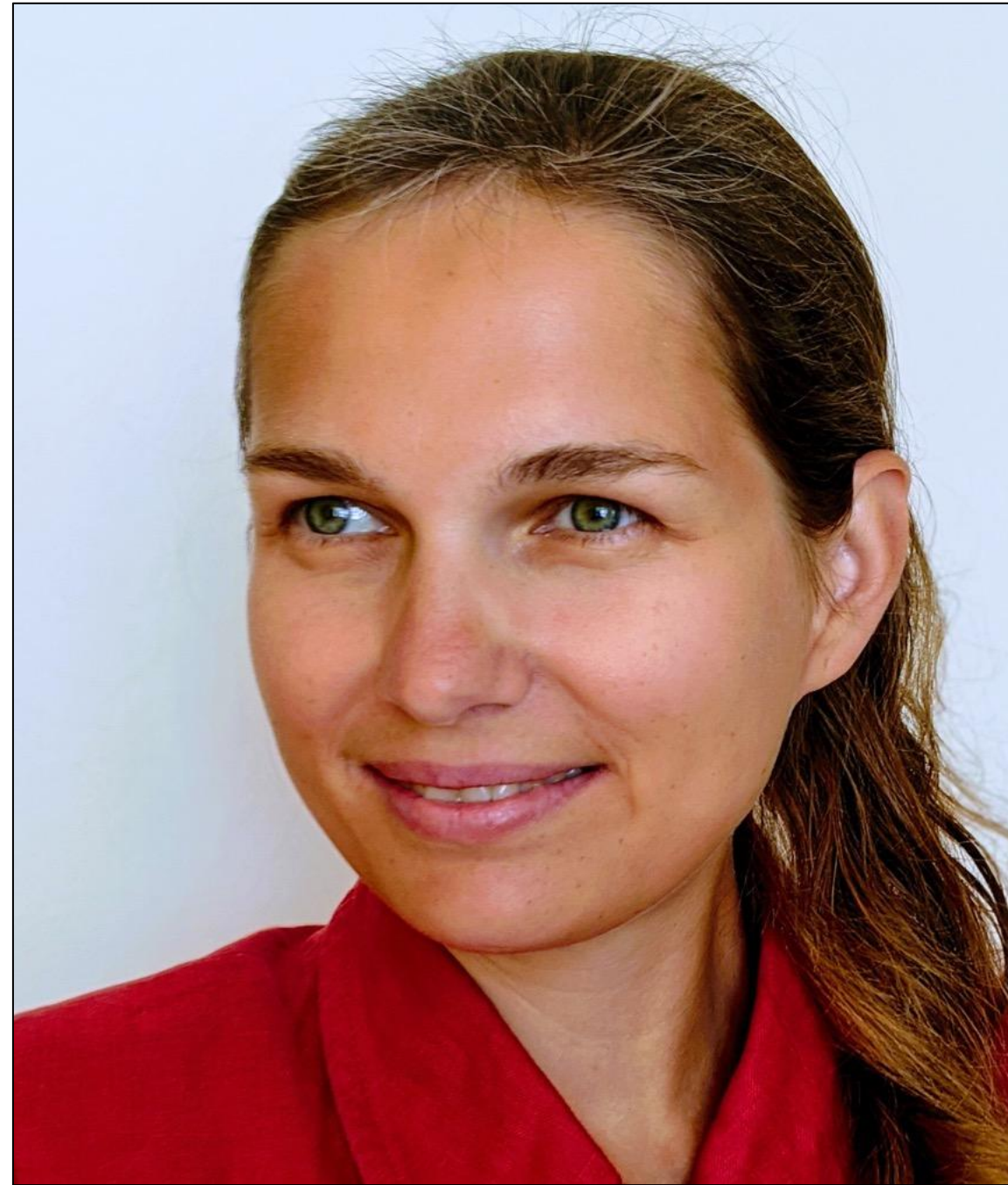


Can “Big Data” and AI help produce better insights to help improve people’s lives?

About us



Dr James Goulding
University of Nottingham



Dr Anya Skatova
University of Bristol



Dr Georgiana Nica-Avram
University of Nottingham

A Highly Multidisciplinary Team:



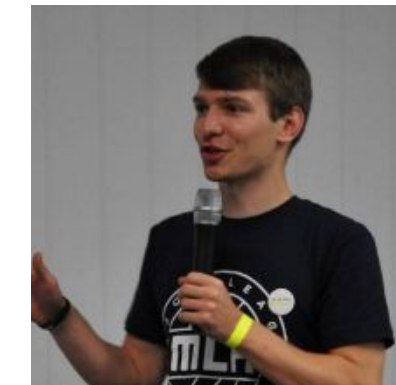
Data Science



Behavioural Science



Geospatial Science



International Development



Consumer Research



Behavioural Psychology



Business Analytics



Data Linkage



Machine Learning



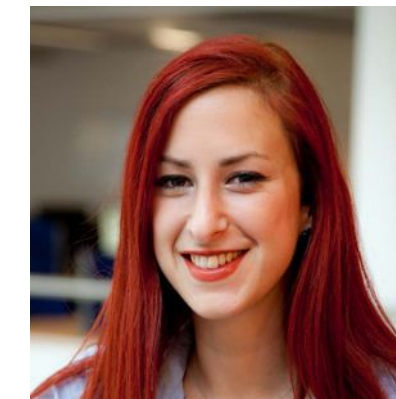
Consumer Analytics



Computer Science



Network Science



Consumer Psychology



Statistics



Qualitative Research



Computer Science



Computer Science



Mathematics & Statistics



Ecology



Computer Science



Behavioural Science

Shopping data: Examples of projects

	Individual level	Aggregated
Standalone	<p><u>Diet and Lifestyle</u>: Lunch time calorie consumption; Context driven choice in food and alcohol</p>	<p><u>Mental Health</u> - understanding seasonal depression/anxiety</p> <p><u>Transport and Mobility</u> - via CDR data</p>
Linked	<p><u>Ovarian Cancer</u>: Early detection of symptoms</p> <p><u>Sustainability</u>: Psychology of plastic bags' consumption</p> <p><u>Consumer Behaviour</u>: Propensity to buy on promotions</p> <p><u>COVID</u>: Individual level validation of COVID-related symptoms</p>	<p><u>COVID</u>: Predicting local outbreaks (CIVIC)</p> <p><u>OTC medication</u>: Seasonal variation in mood and OTC; OTC medication and deprivation;</p> <p><u>Food Poverty</u>: Identifying UK food insecurity</p> <p><u>Period Poverty</u>: Menstrual pain and deprivation</p> <p><u>Laxatives mis-use</u>: Assessing policy change</p> <p><u>Chronic pain</u>: Purchase of painkillers and part time working</p> <p><u>Diabetes</u>: Links between risk and sugar</p>

Modelling and mapping food-insecurity

Food-insecurity is a **widely reported problem in the UK:**

→ Yet measurement is almost *non-existent*.

→ The few exceptions are surveys that are one-off, expensive, geographically sparse, and soon out of date.

→ The consequence is that existing data is irrelevant to local authorities trying to help **communities in need.**

Round	Surveyer	Dates conducted	Sample Size	Participants	Country
1	YouGov	25th-26th March 2020	2070	Adults (18+)	Great Britain
2	YouGov	7th-9th April 2020	4343	Adults (18+)	Great Britain
3	YouGov	24th-29th April 2020	2284	Adults in households with children	United Kingdom
4	YouGov	14th-17th May 2020	4352	Adults (18+)	United Kingdom
5	YouGov	6th-8th July 2020	4350	Adults (18+)	United Kingdom
6a	Childwise	8th-20th September 2020	1064	Children aged 7-17 (Year 2+)	United Kingdom
6b	YouGov	24th August-1st September 2020	10845	Adults (18+)	United Kingdom
7a	Childwise	22nd Jan-2nd February 2021	1308	Children aged 7-17 (Year 2+)	United Kingdom
7b	YouGov	29th Jan-2nd February 2021	4231	Adults (18+)	United Kingdom
8	YouGov	4th-9th August 2021	6490	Adults (18+)	United Kingdom
9	YouGov	18th-20th January 2022	4186	Adults (18+)	United Kingdom
10	YouGov	22nd-29th April 2022	10,670	Adults (18+)	United Kingdom

The Food Foundation
All data is under an open-source policy. For further information please contact: shona.gould@foodfoundation.org.uk

Coronavirus: Panic buyers strip shelves as England prepares for lockdown

Despite stores remaining open and assurances that there is enough for everyone, some shoppers appear to be panic buying.

Tuesday 3 November 2020 15:16 UK

COVID-19 CORONAVIRUS



Marcus Rashford petition to end child food poverty signed by over 1m people

It comes as pressure builds on government to U-turn on free school meals



▲ Marcus Rashford and his mother, Melanie, visit FareShare in Greater Manchester on 22 October. Photograph: Mark Waugh/AP

Government outsourcing of food parcels forced clinically vulnerable to shield without food



The Government's food box scheme was supposed to be a lifeline for the clinically vulnerable. They were told to stay inside and often could not get deliveries from supermarkets. But our research suggests the scheme was used to palm off on the vulnerable unwanted, and sometimes unsuitable, supplies from elsewhere in the food chain. And that the Government's preferred contractors, chosen without a tender process, may have reaped substantial profits...



University of
Nottingham

UK | CHINA | MALAYSIA



University of
BRISTOL

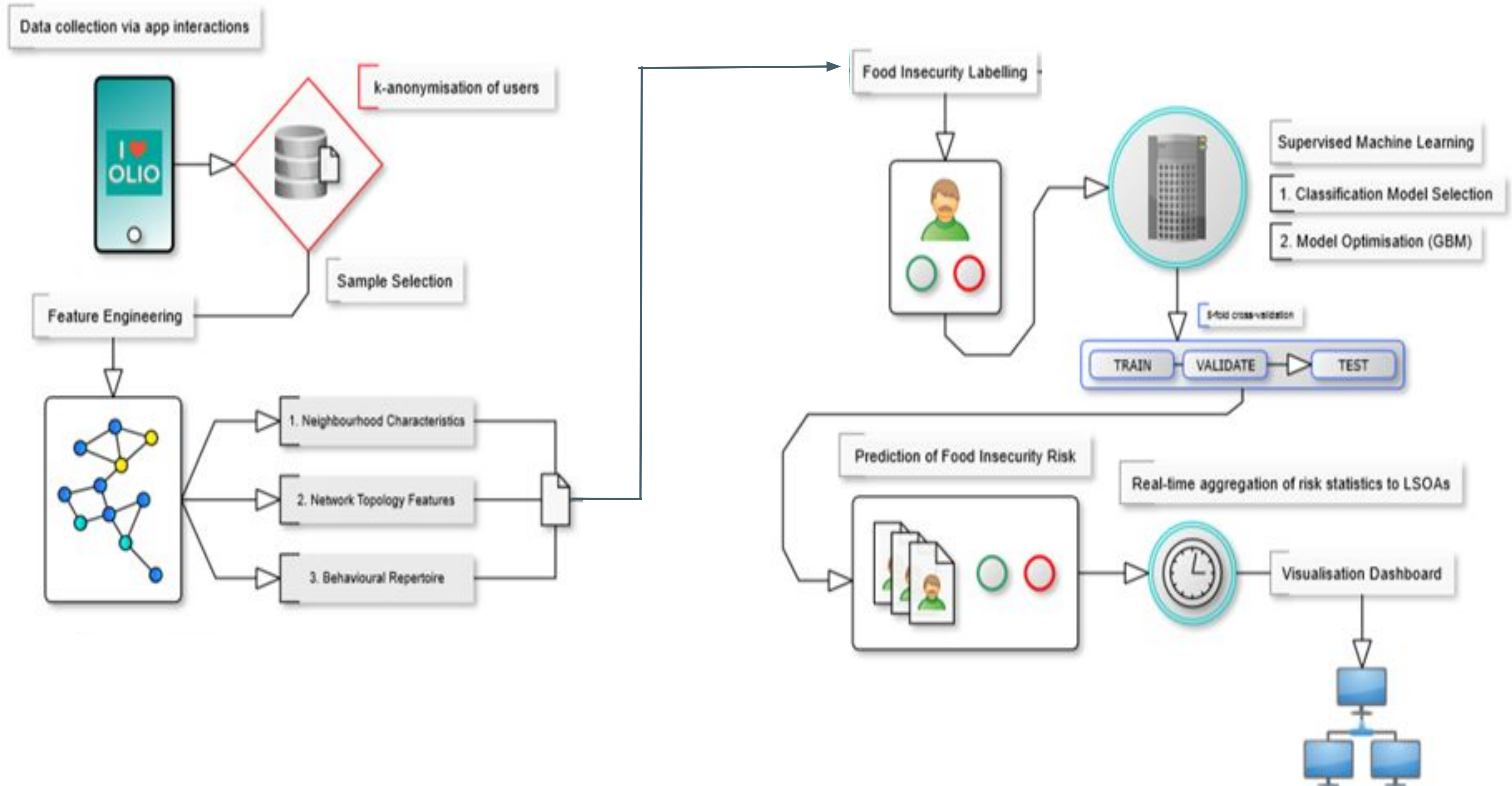
A large, high-resolution image of the Earth as seen from space, showing the Western Hemisphere. The Earth is centered in the frame, with the Atlantic Ocean, North America, and South America visible. The image is set against a dark, starry background of space. A white rectangular border is superimposed over the Earth image, containing the main text and logos.

Levelling-up food insecurity systems

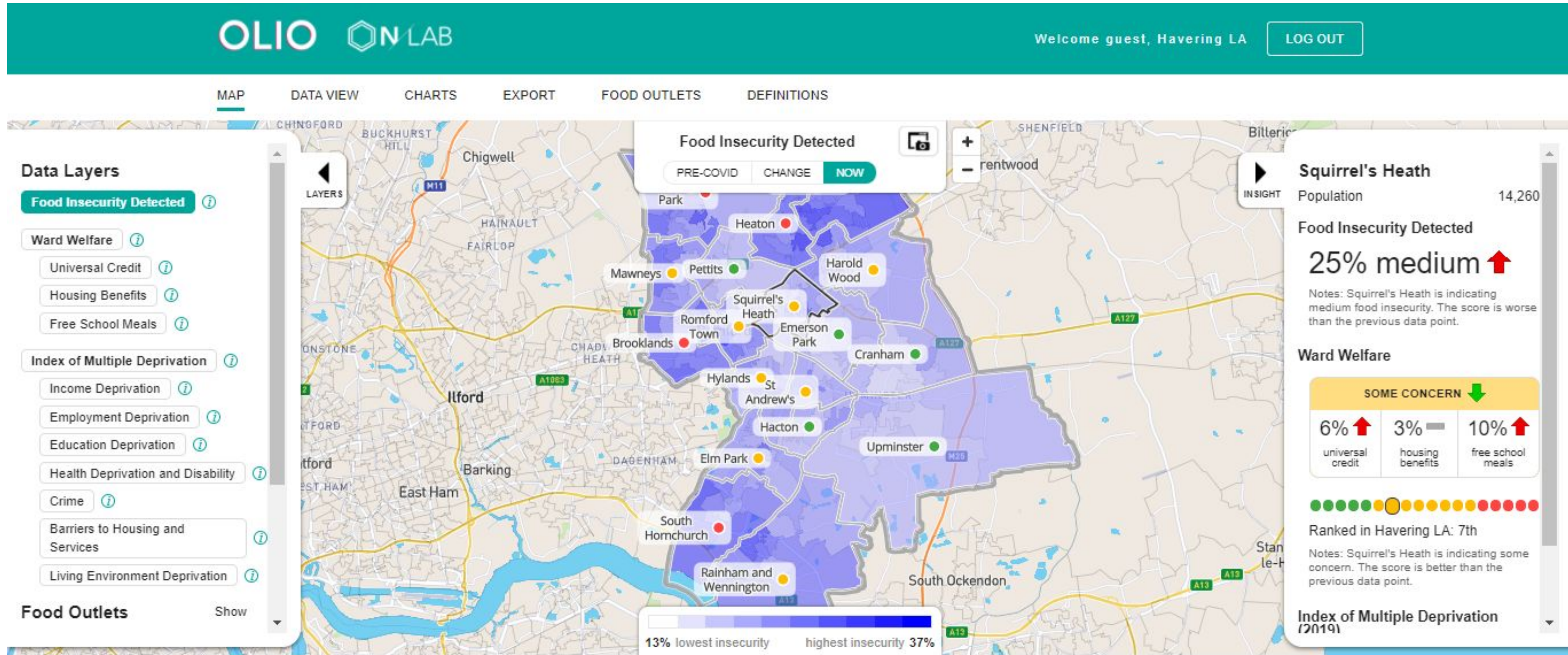
N / LAB

Digital
Footprints

Partnership with Havering Council and London Assembly



Partnership with Havering Council and London Assembly

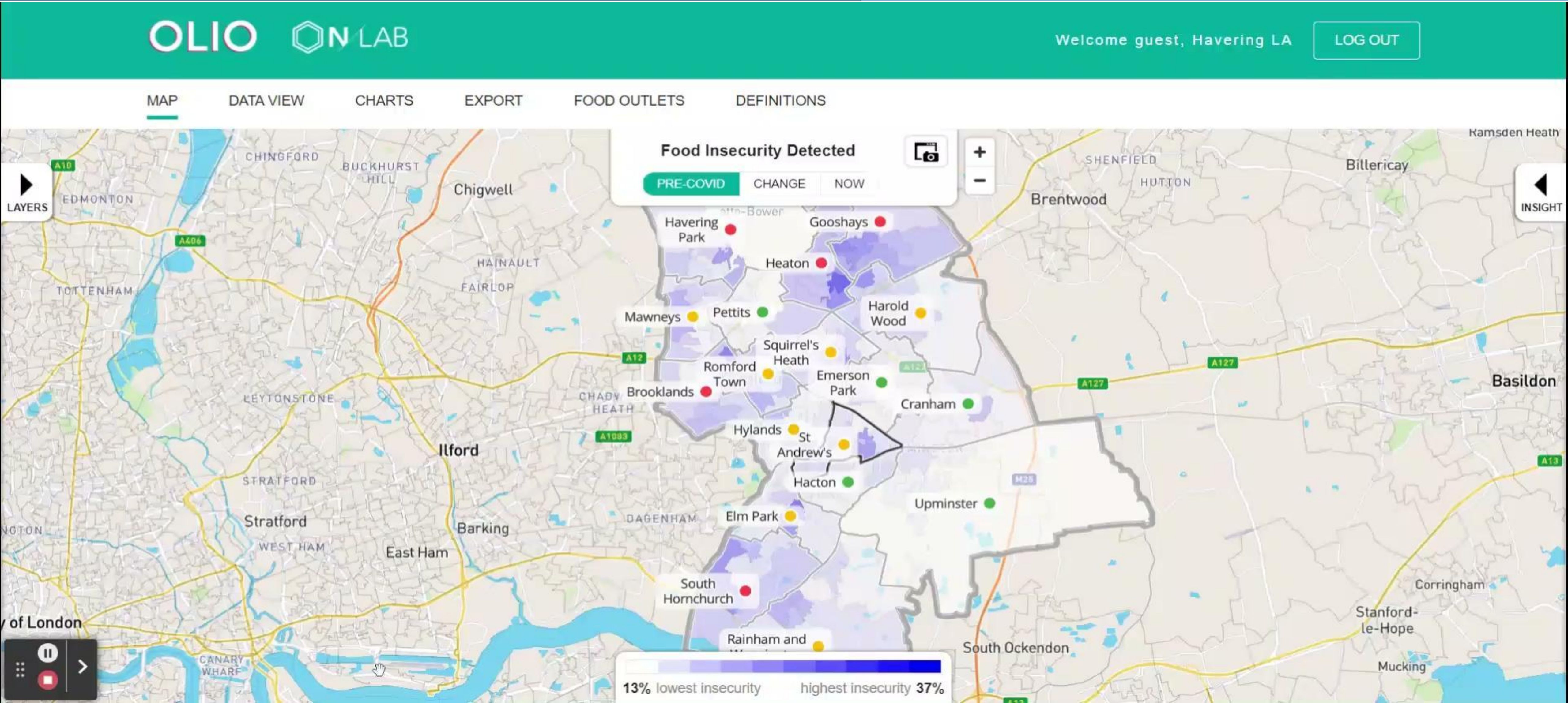


→ Produce food-security prevalence estimates at local area level.

→ Use interface to measure effectiveness of food interventions in Havering.

→ Further opportunities to integrate nutritional and health information.

Partnership with Havering Council and London Assembly



Reforming the understanding of food-insecurity

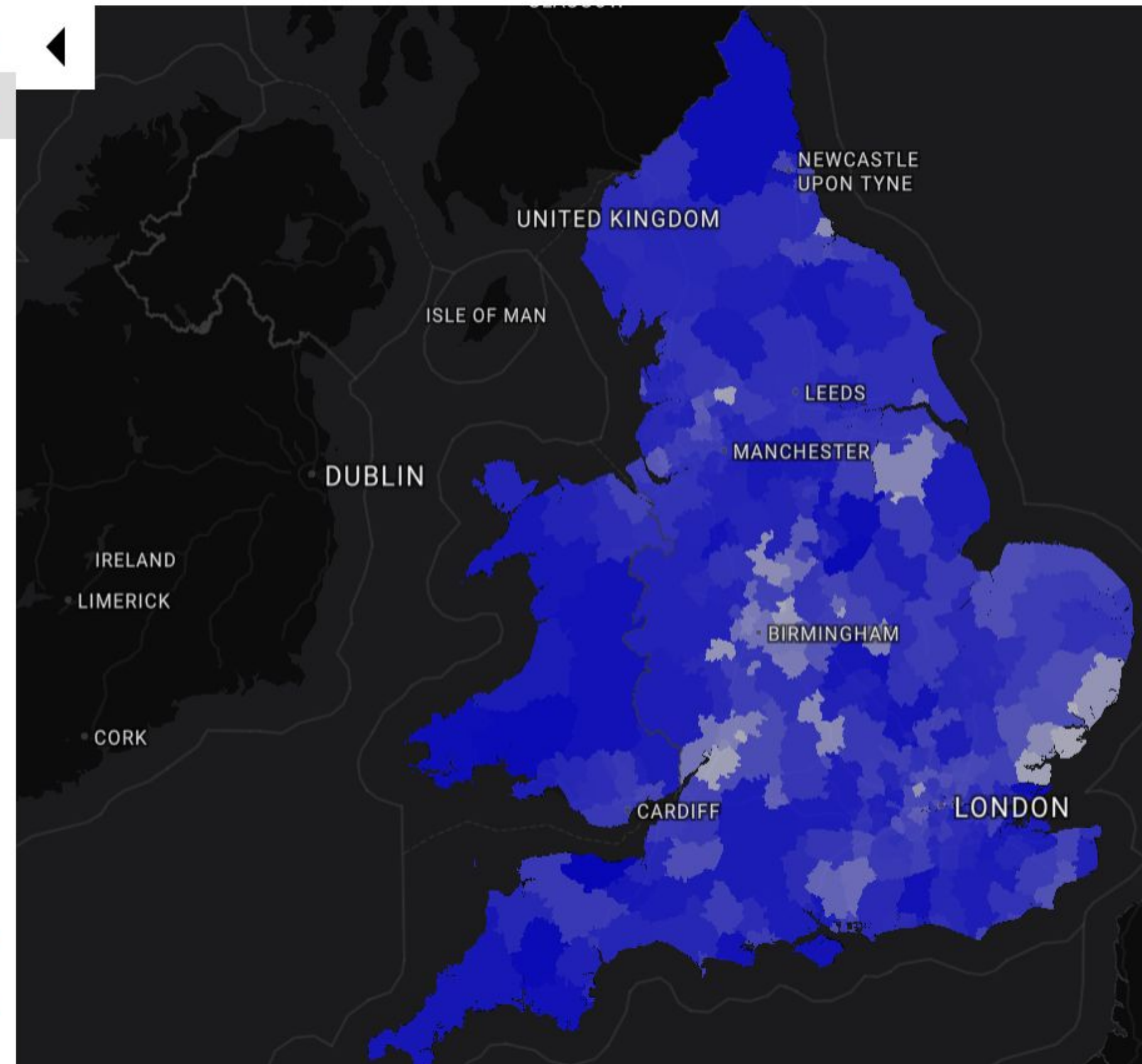


Availability Access
Utilization **Stability** **Agency** **Sustainability**



UK Nutrition Security Map 2020-4

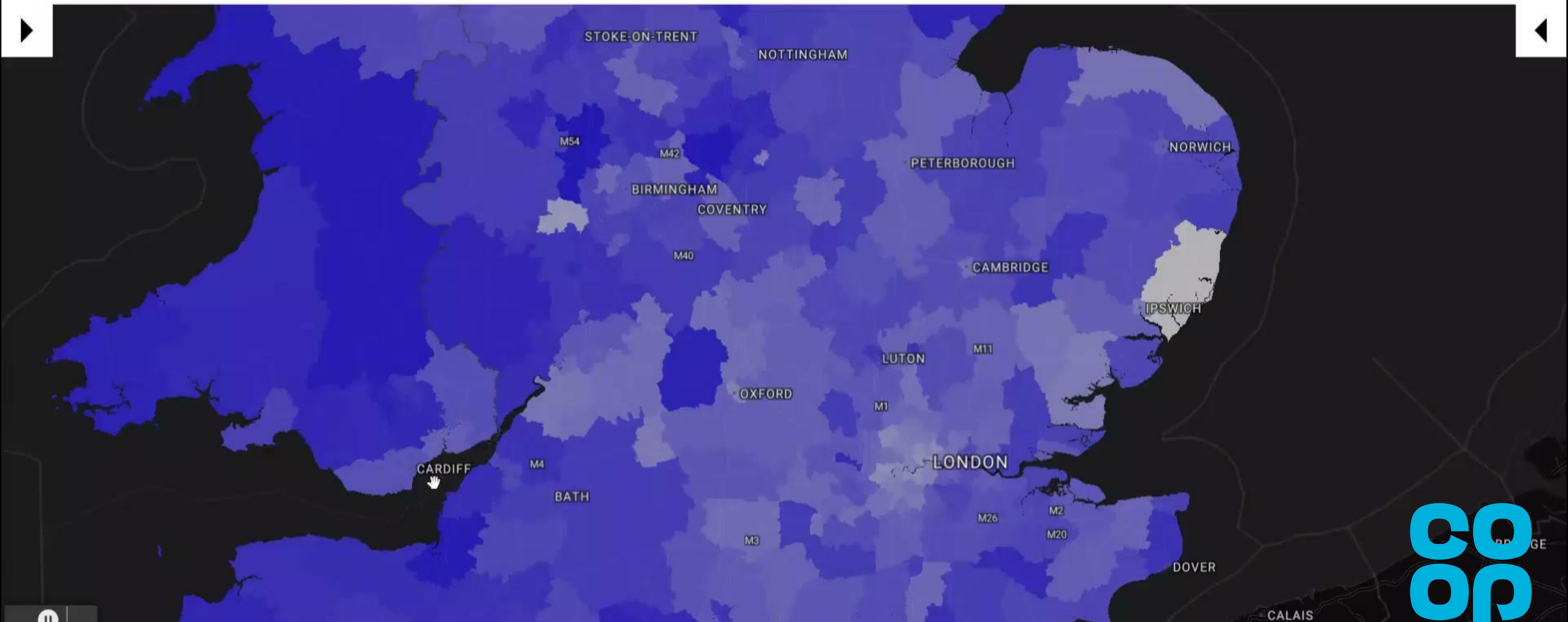
- CONSUMER BEHAVIOUR +
- BASKET COMPOSITION -
- Fruit & Vegetables
- Grains
- Red Meat
- Poultry
- Fish
- Dairy
- Eggs
- Fats & Oils
- Canned
- Sweets
- Cigarettes
- Ready-Made
- Sauces & Soups
- Tea & Coffee
- Soft Drinks
- Beer, Lager & Cider
- Wine
- Spirits
- NUTRITION DATA +
- DEPRIVATION STATISTICS +



Reforming the understanding of food-insecurity

UK Nutrition Security Map

[logout](#)






University of
Nottingham

UK | CHINA | MALAYSIA



University of
BRISTOL

A large, high-resolution image of the Earth as seen from space, showing the curvature of the planet and the blue oceans. The image is centered in the background of the slide.

Levelling-up health early warning systems

N/LAB

Digital
Footprints

Misuse of Laxatives

Recent intervention made it more difficult to buy stimulate laxatives - however it is not clear whether this made any different to overall sales, including for misuse. We can evaluate such policies through analysis of sales of products with laxatives.

Chronic pain

Purchases of painkillers are higher in the areas where there are more part time workers.

Period pain

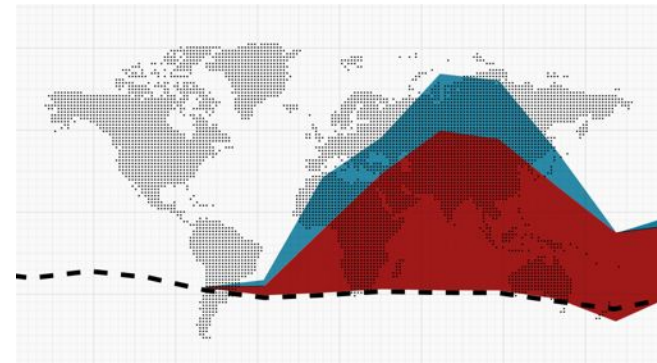
Purchases of painkillers bought simultaneously with period products can be used to estimate prevalence of period pain. $\frac{1}{3}$ of those who bought menstrual products, simultaneously also bought pain products.

Respiratory illness

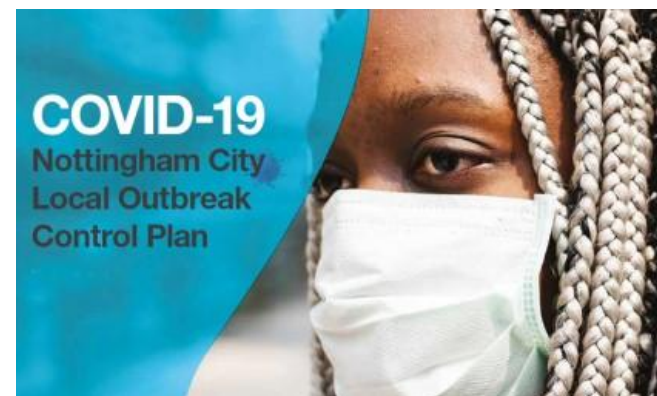
Purchases of cough medication and other flu remedies can predict spikes in respiratory deaths.

The Research Questions:

CIVIC



- Could we have *improve* **estimation** of unrecorded cases of COVID that occurred in the UK using digital footprint data?



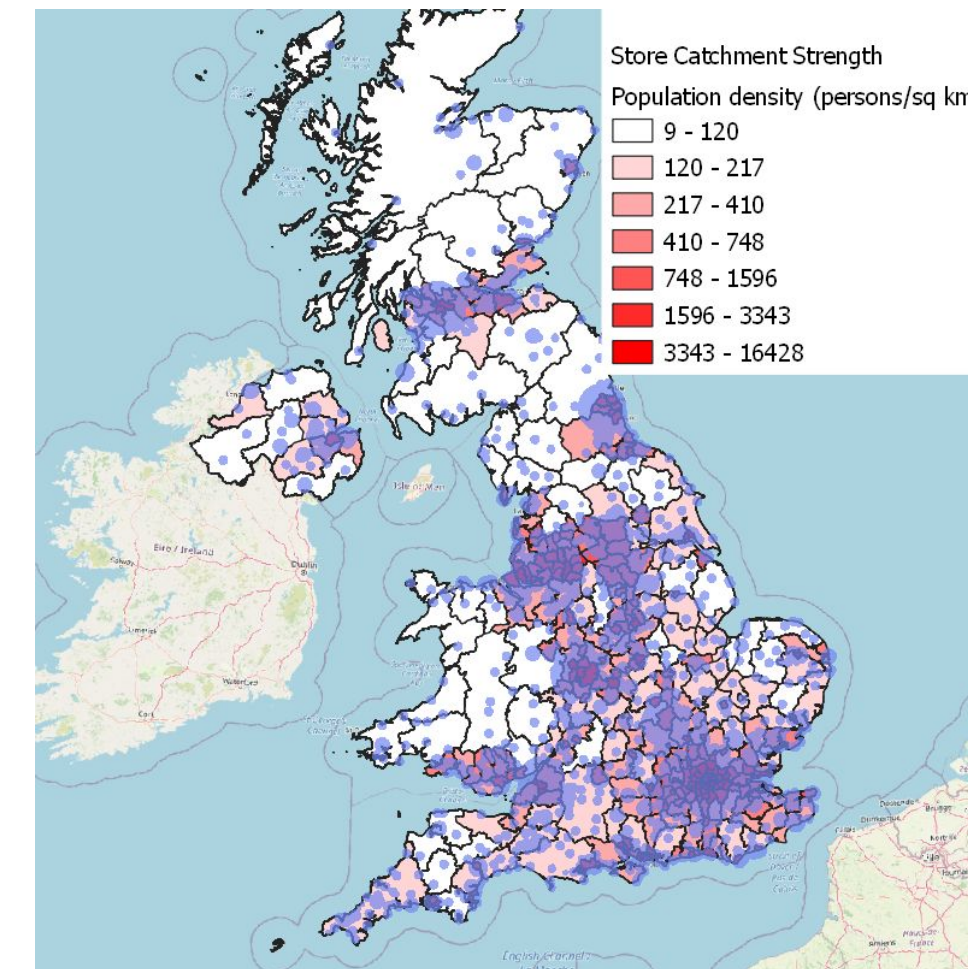
- Can we identify **antecedents** of COVID deaths in that footprint data **ahead of time?**
*(for early-warning systems at scale; **without reliance on self-reporting apps**)*



- What insight can such datasets give us on the impact to **vulnerable**, potentially hidden, communities (e.g. food poverty, BAME), to help long-term intervention strategies.

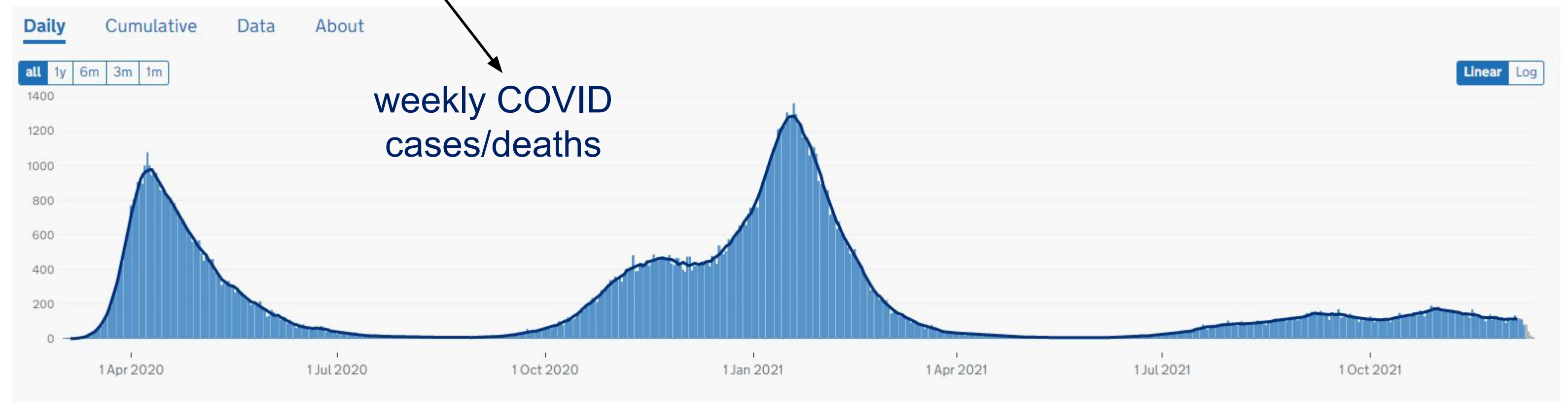
The CIVIC Platform

The CIVIC Platform:



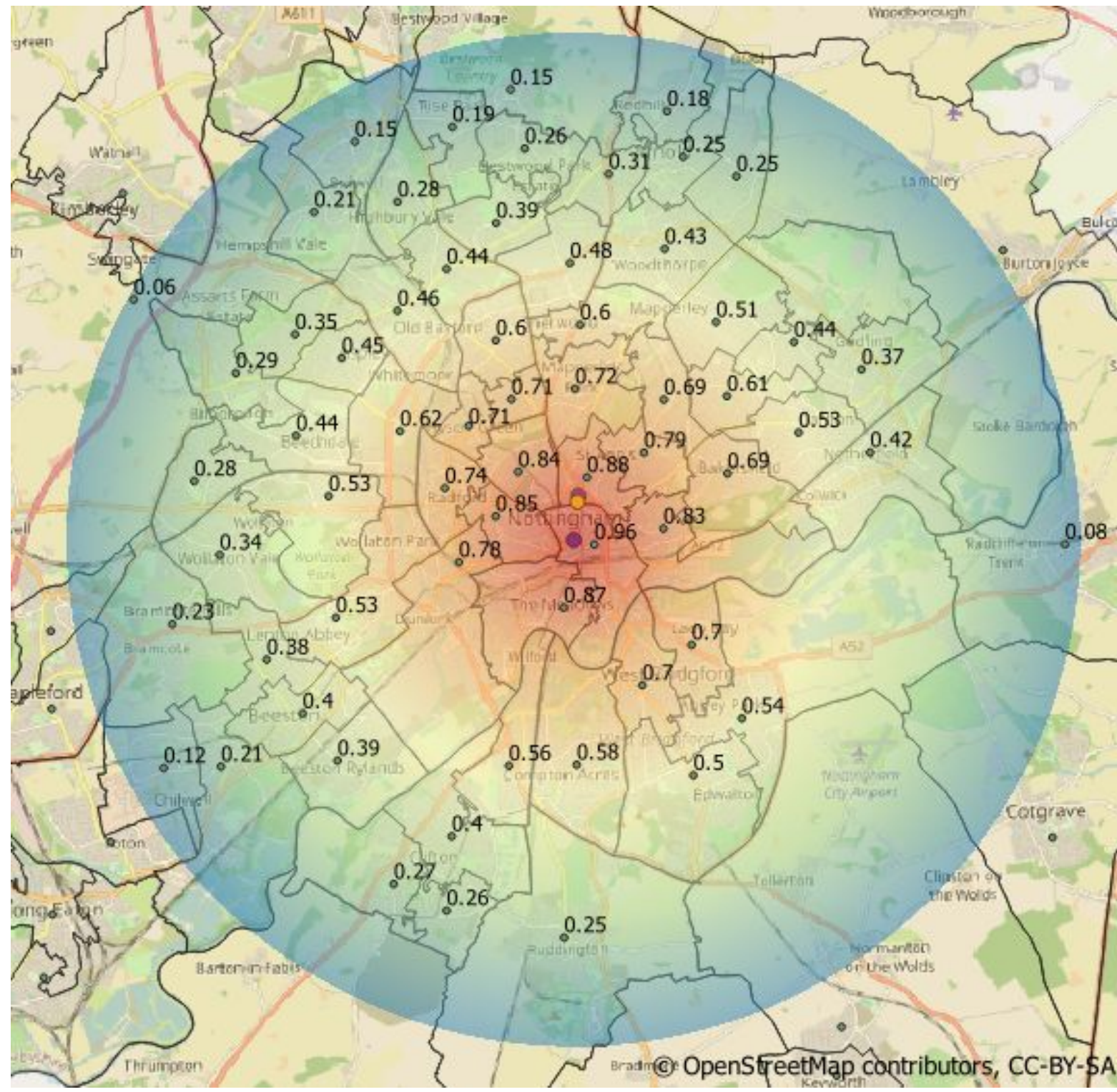
BASIC STATS:

- Data using in models is aggregated to areas (2016 to 2021) but reflects **2,155,300,393** sale units in England (*not including online*)
- For example, 14,380,021 cough medicine sales

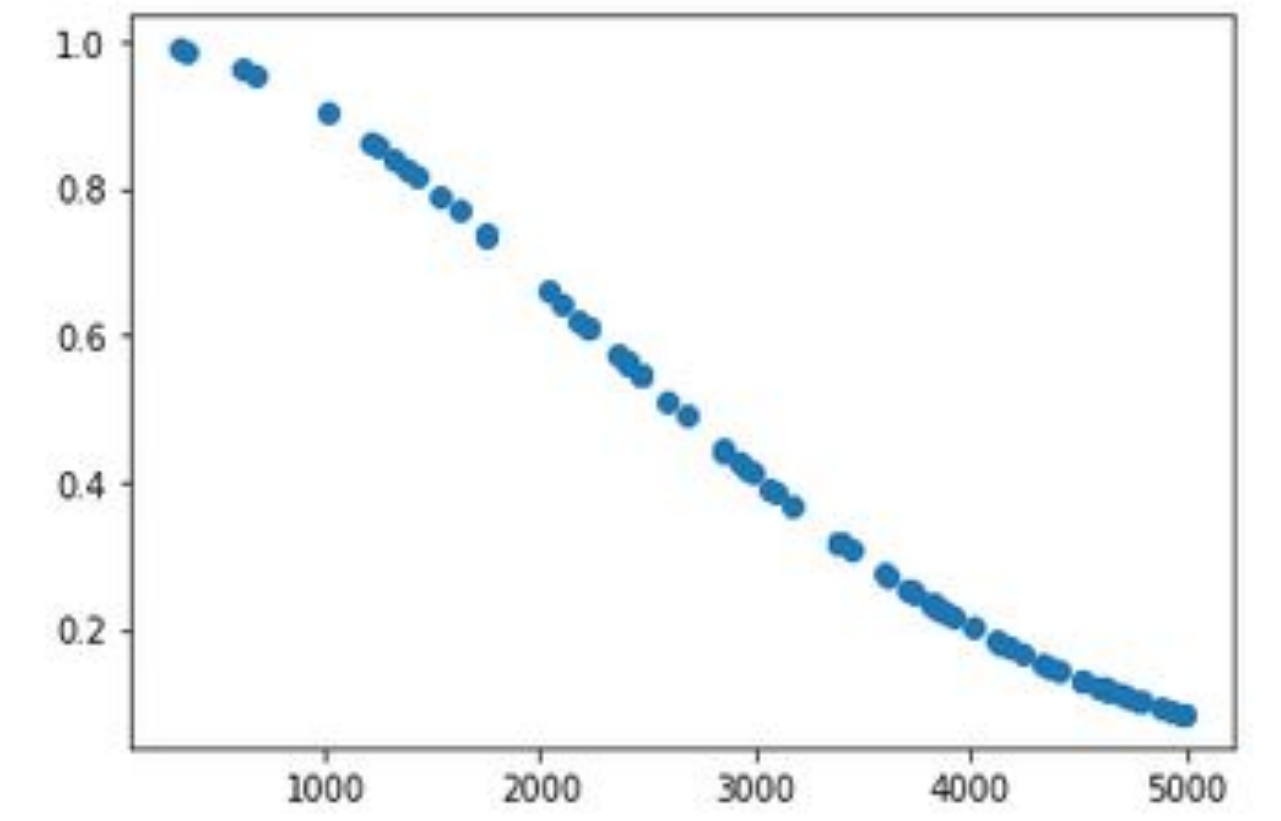
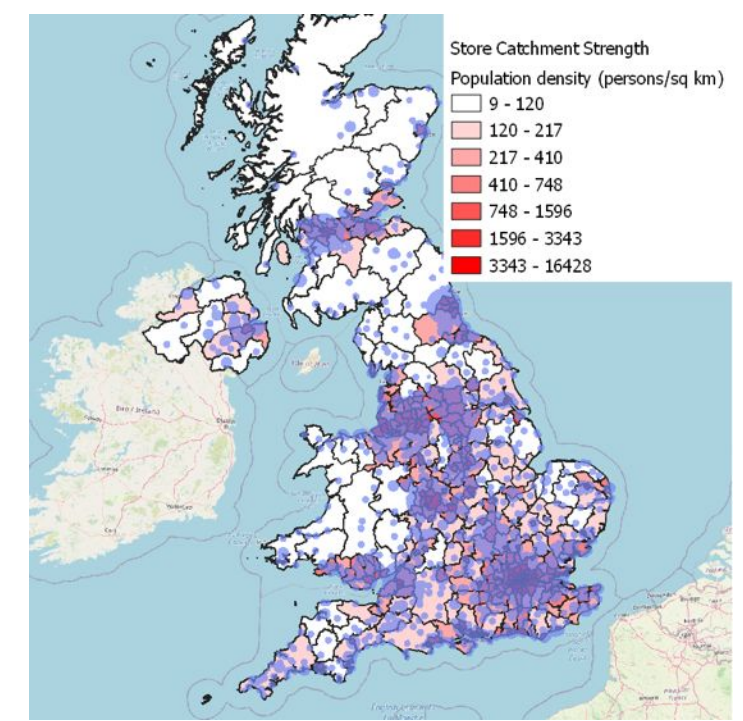
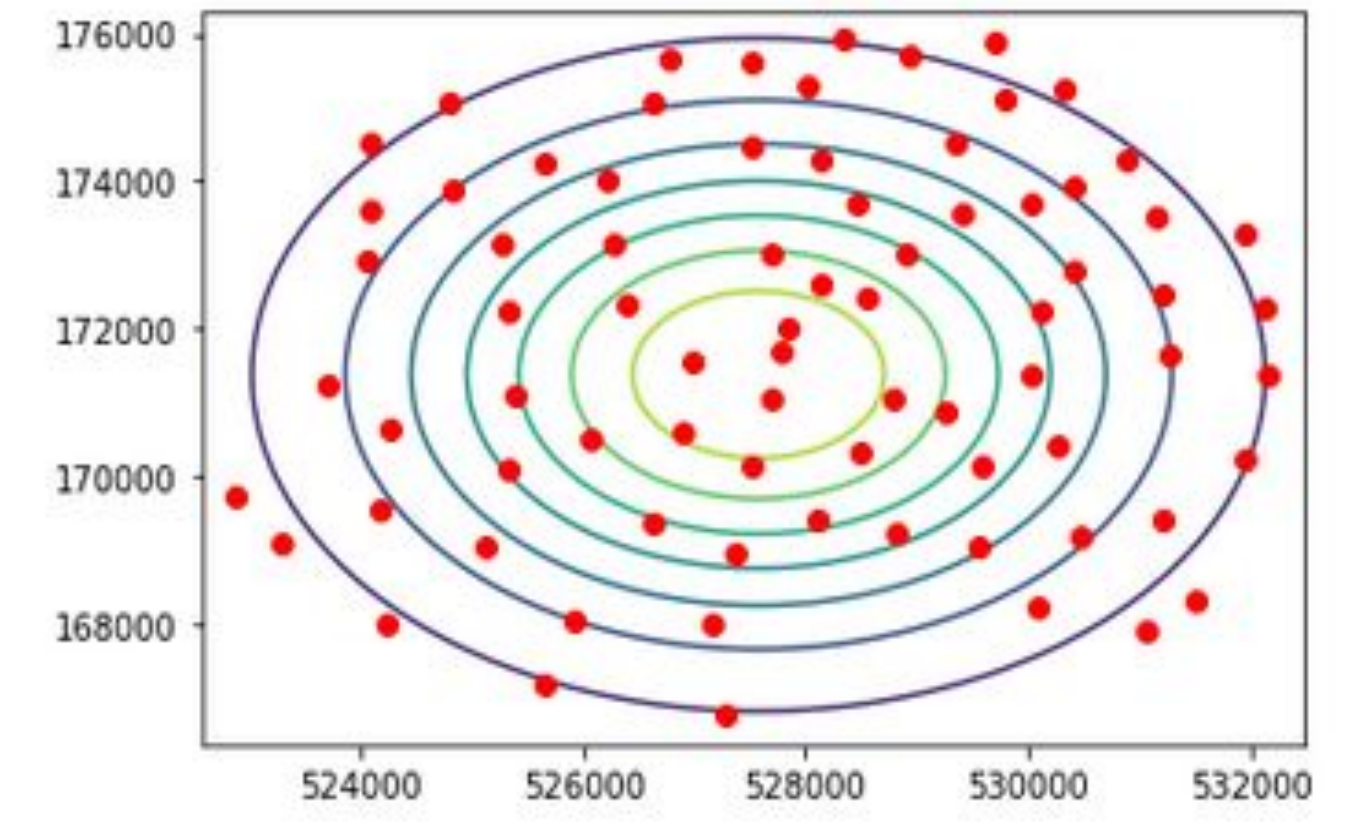
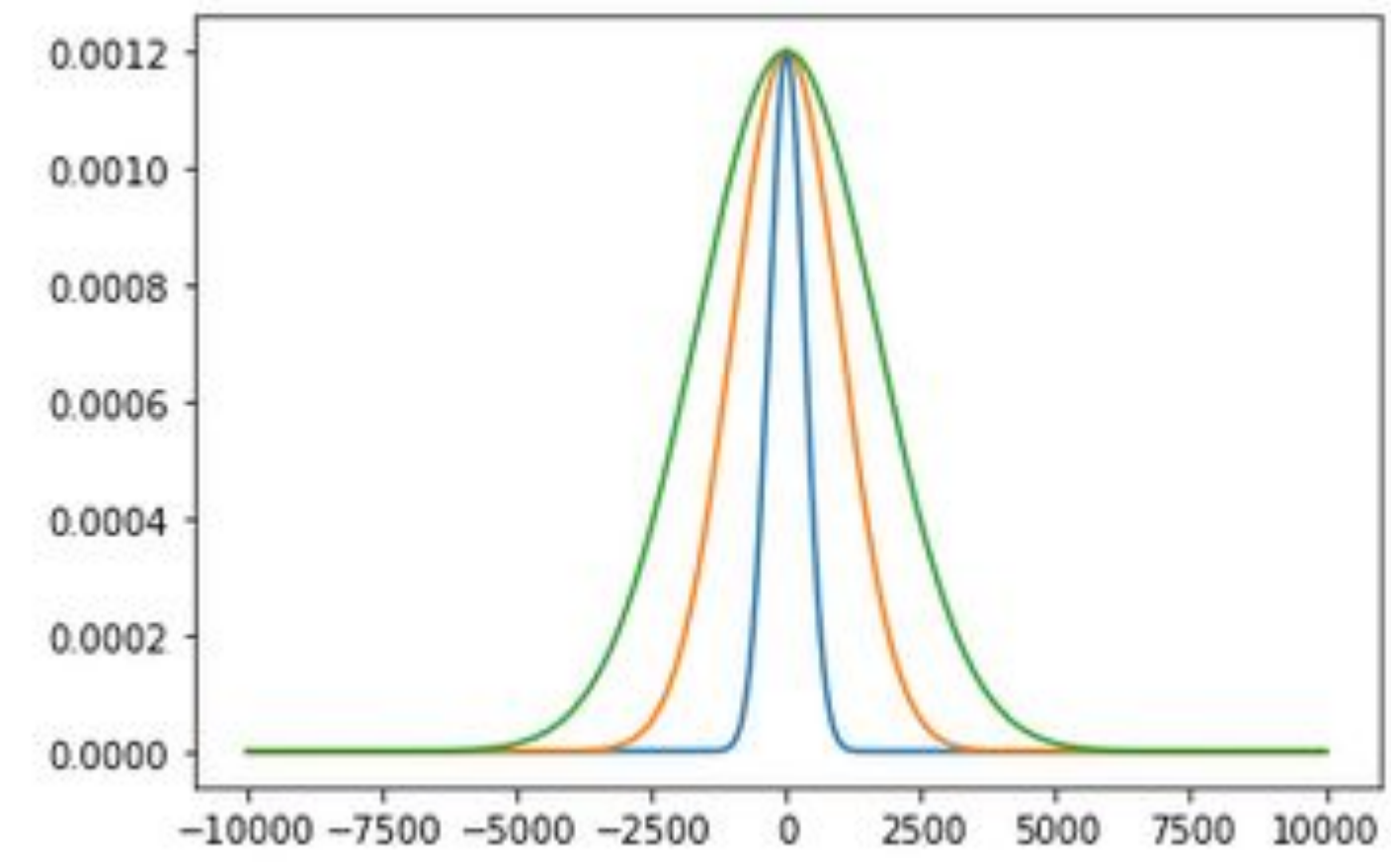


Modelling Catchment Areas

Using a normal distribution to assign 'probability' that the sale came from an individual in that MSOA



*Middle Layer Super Output Area



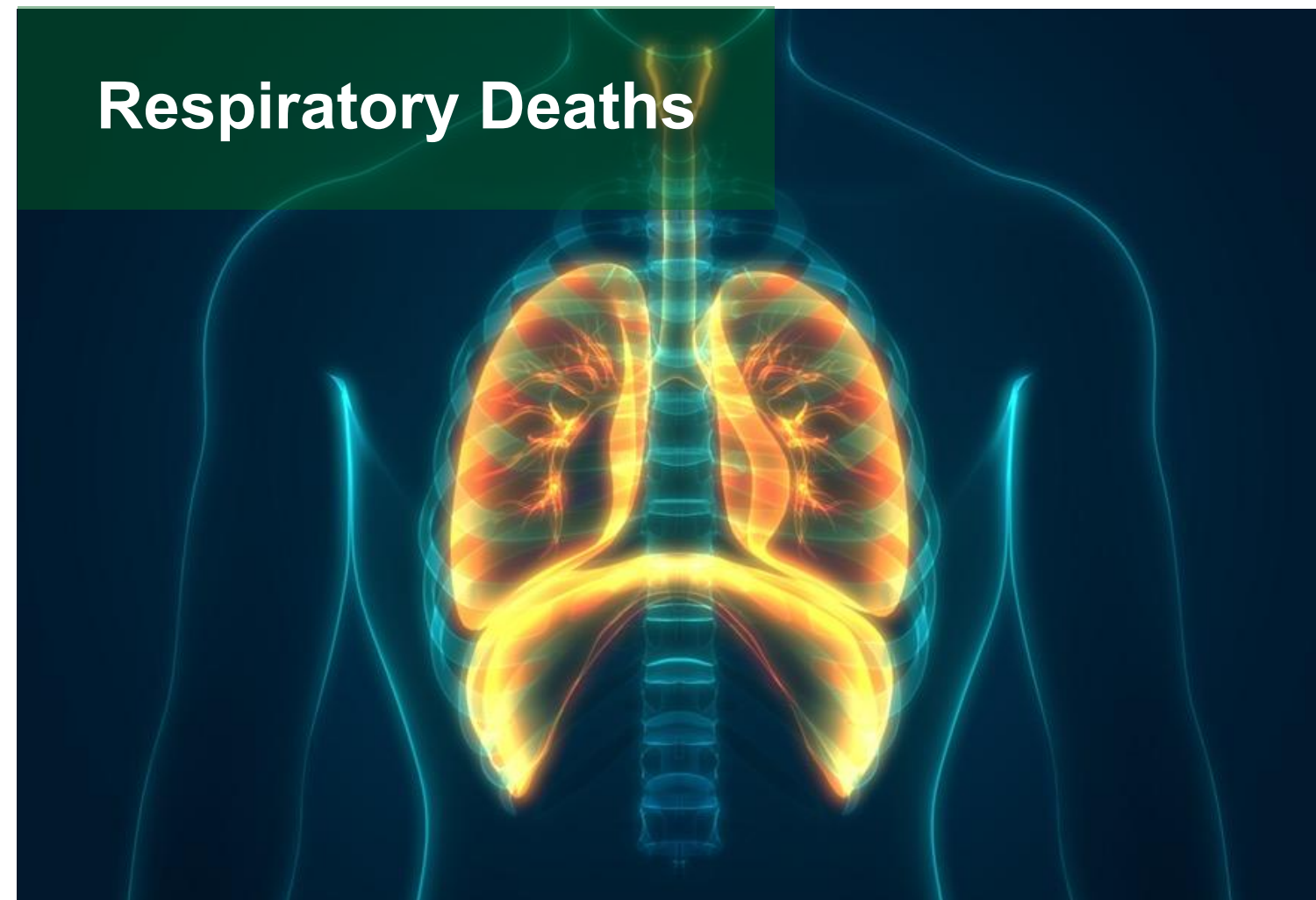
You can't use sales alone...



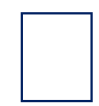
*Not only does behaviour gradually change but edicts caused huge “shocks” in product purchases....
...which change correlations between sales and the disease,*



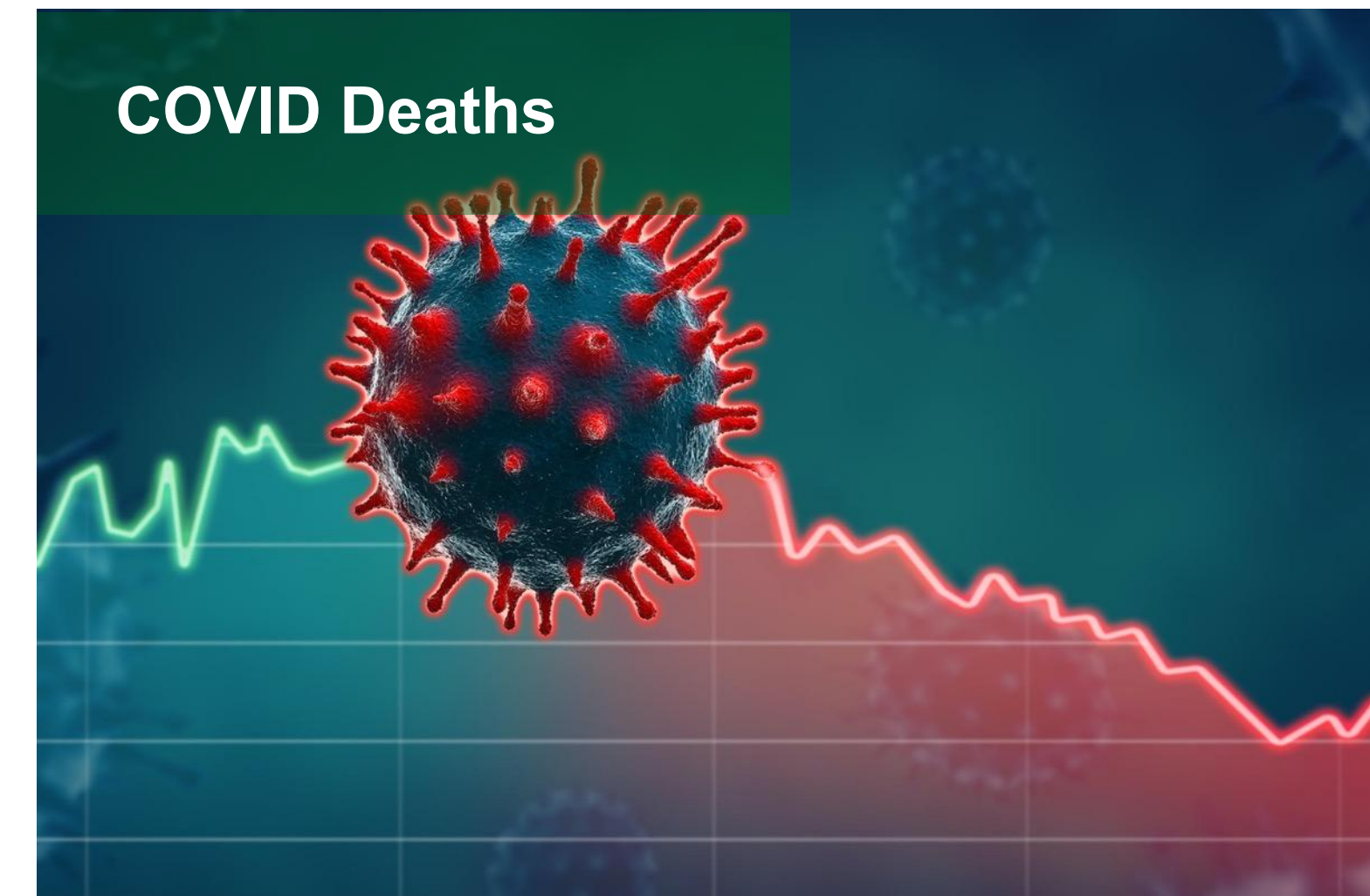
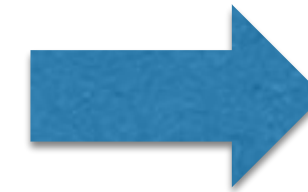
Respiratory Illnesses First!



18th March 2016



27th March 2020



28th March 2020



1st Jul 2021

The PADRUS* model

Model type: Random Forest Regressor.

(Optimized using a time series cross-validation grid-search on training data to prevent over-fitting).

Target: Predict weekly respiratory deaths 17 days in advance for each of the 314 LTLA (Lower Tier Local Authority) areas in England from 18th March 2016 to 27th March 2020.

Inputs: 56 features

Training Datapoints: 45,844

Testing Datapoints: 20,410

Results from predicting <u>on test data</u> :		
Mean Absolute Error	Root Squared Mean Error	R ²
2.39	3.42	0.78



***Predicting amount of deaths from respiratory disease using sales data**

Features for our model

Sales features

- lta_week_sales_17
- decongestant_17
- throat_17
- cough_dry_17
- cough_all_17
- lta_week_sales_24
- decongestant_24
- throat_24
- cough_dry_24
- cough_all_24
- decongestant_lsr
- throat_lr,
- cough_dry_lsr
- cough_all_lsr
- decongestant_lsm
- throat_lsm
- cough_dry_lsm
- cough_all_lsm

Seasonal features

- average_rainfall
- total_rainfall
- min_temp
- average_temp
- max_temp
- week_num

Land Use features

- pct_community
- pct_industrial
- pct_residential
- pct_transport
- pct_agriculture
- pct_natural
- pct_recreation

Deprivation features

- liv_env_score
- crime_score
- housing_score
- imd_rank
- imd_score
- imd_extent
- imd_concentration

Demographics

- pct_male
- pct_female
- pct_non_white
- pct_lone_parent
- pct_other_children

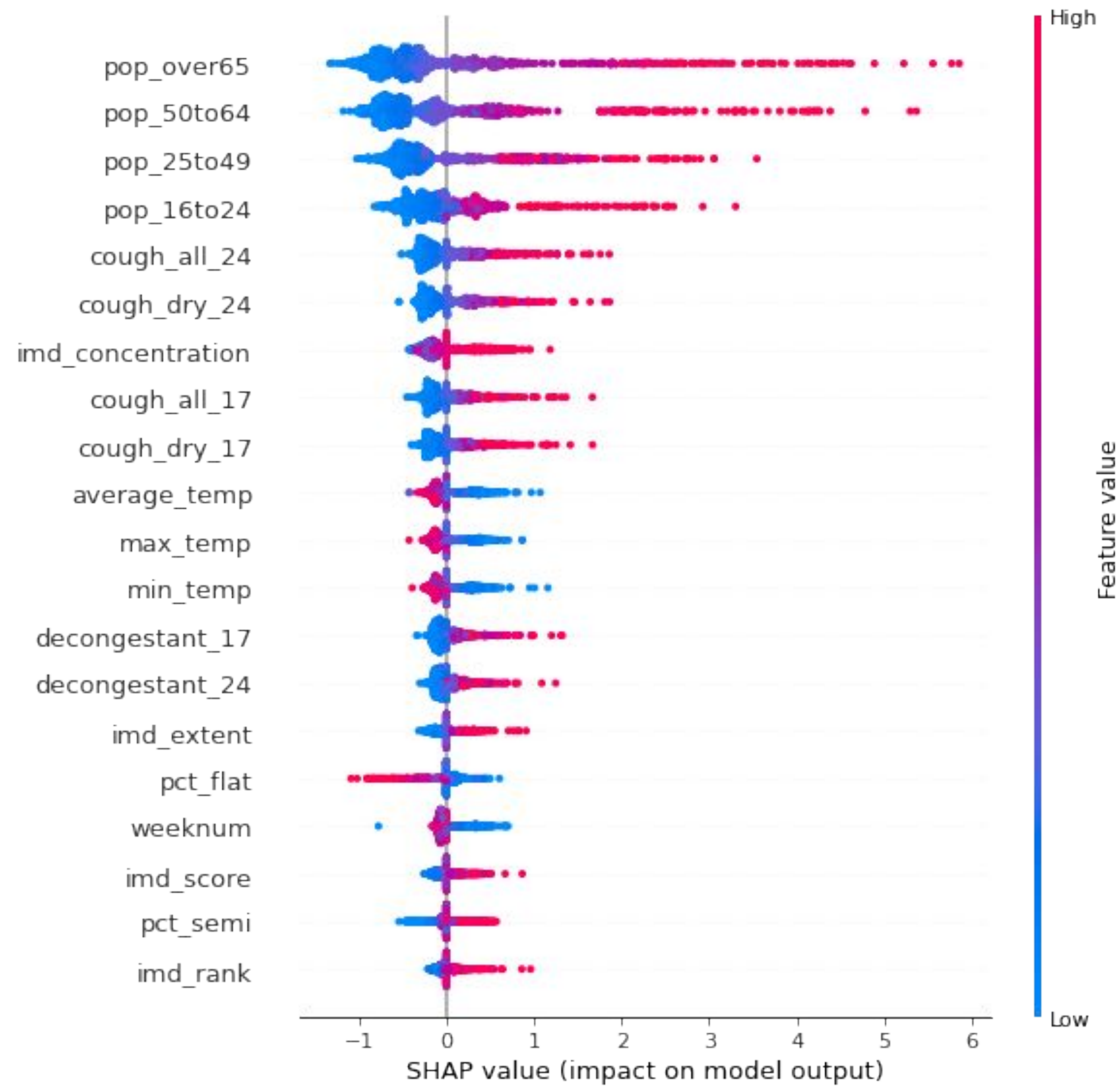
Population features

- pop_density
- pop_16to24
- pop_25to49
- pop_50to64
- pop_over65

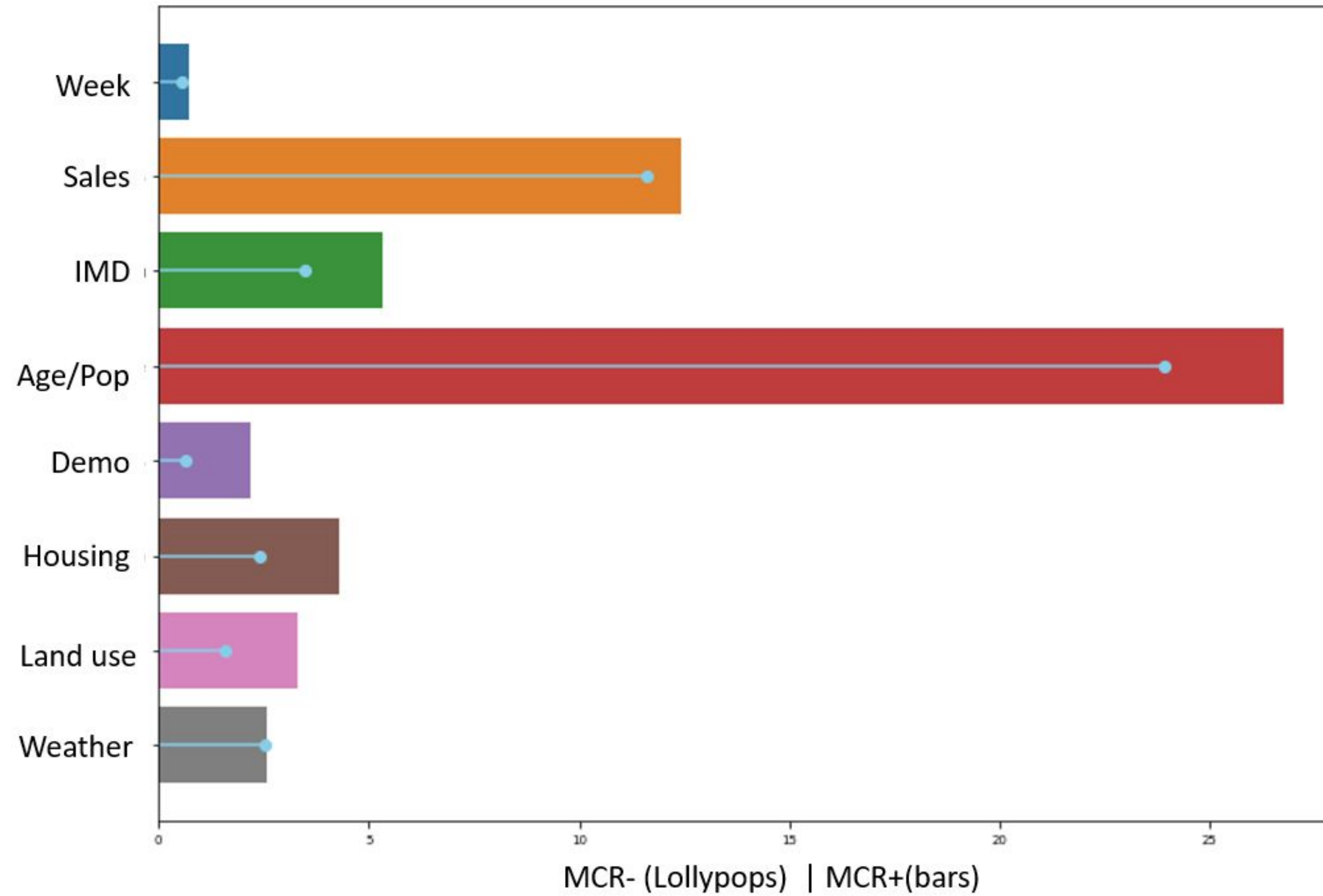
Housing Features

- pct_pre1919
- pct_pre1940
- pct_pre1973
- pct_pre1983
- pct_detached
- pct_semi,
- pct_terraced
- pct_flat

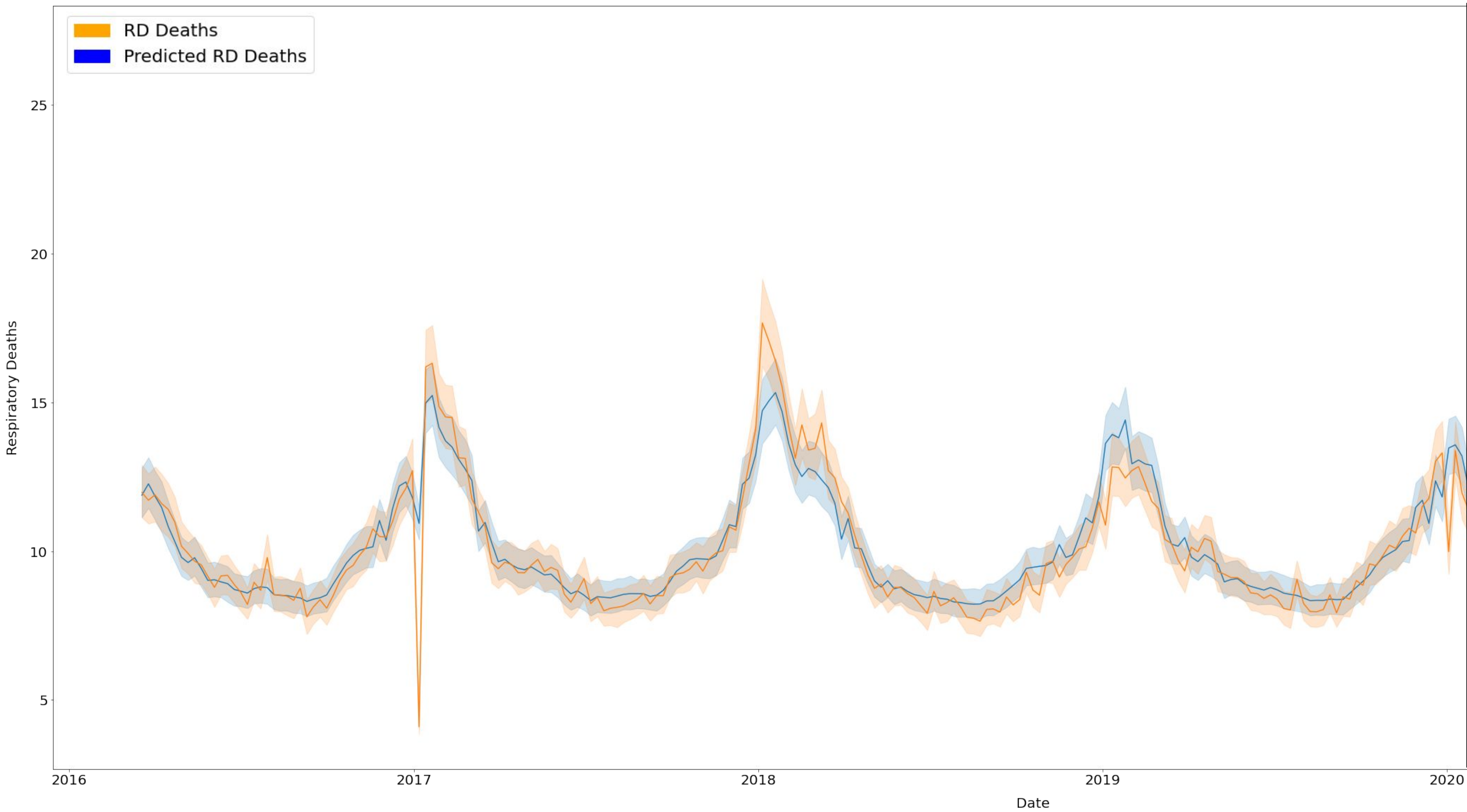
Variable Importance Analysis - Shapley



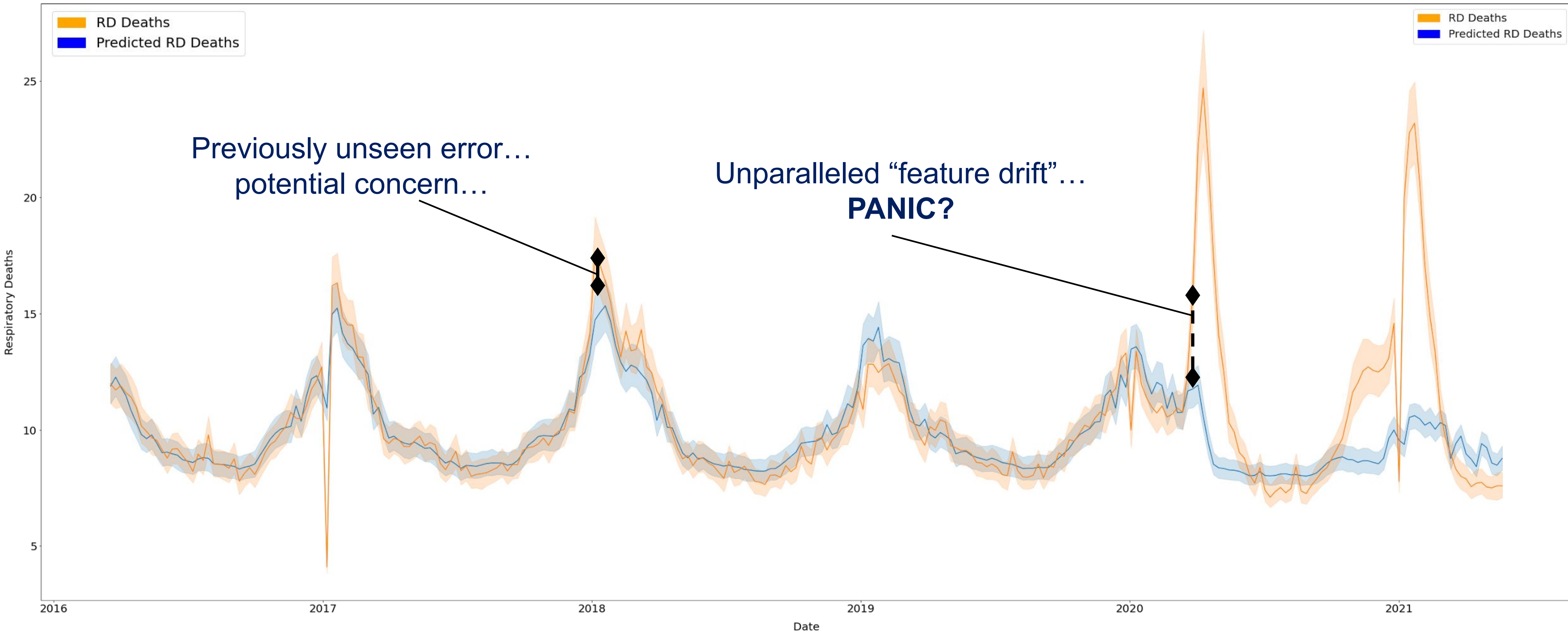
Variable Importance Analysis - MCR



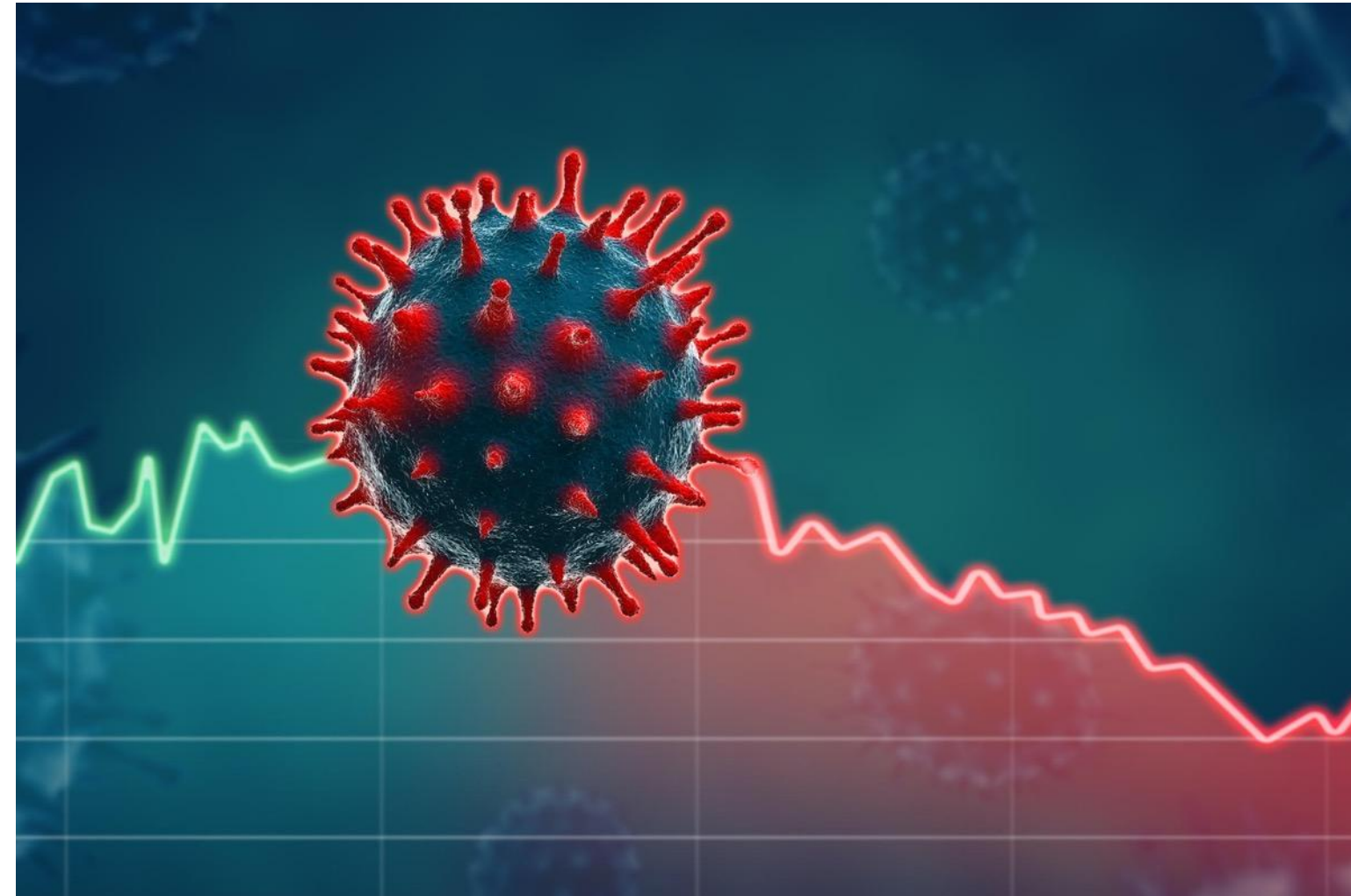
Forecasting Deaths WITH sales



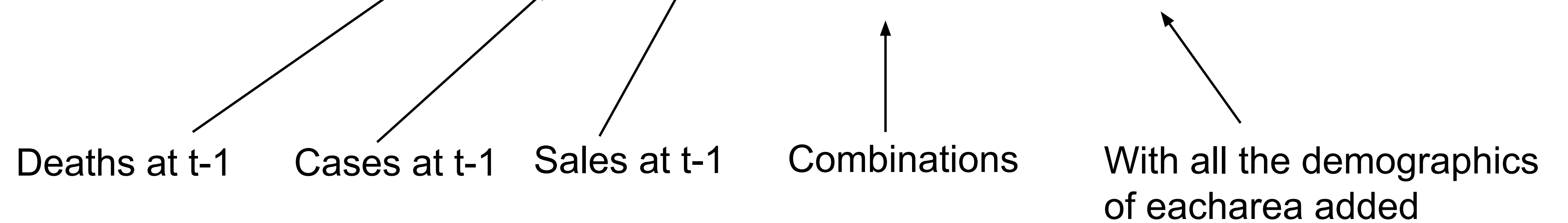
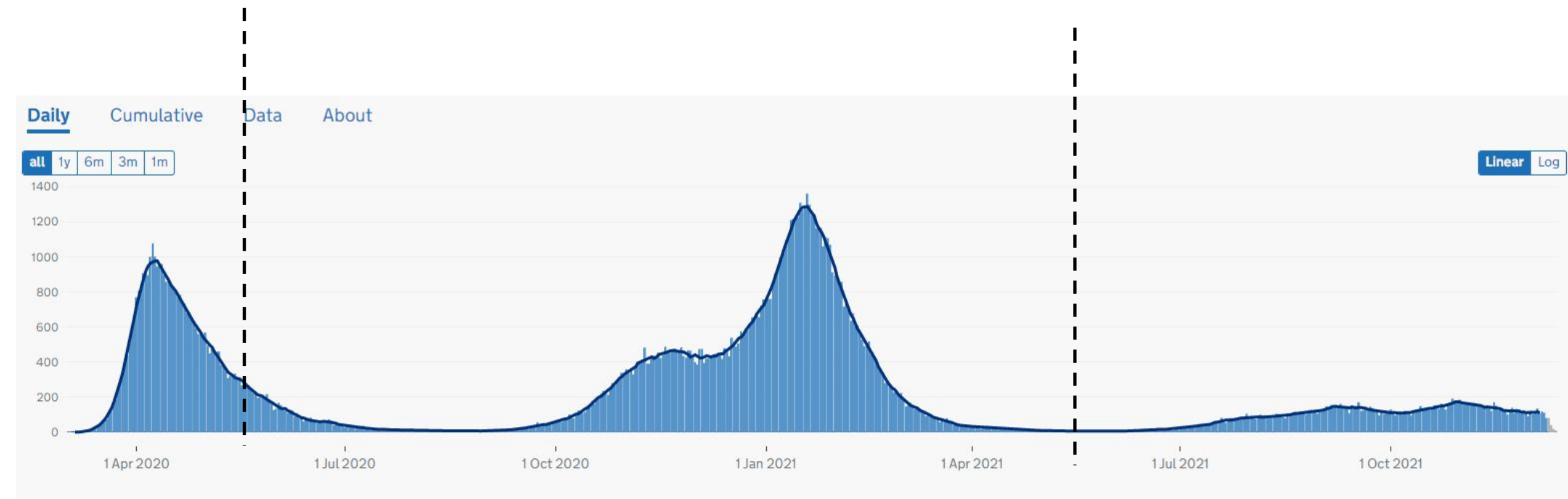
Forecasting Deaths WITH sales



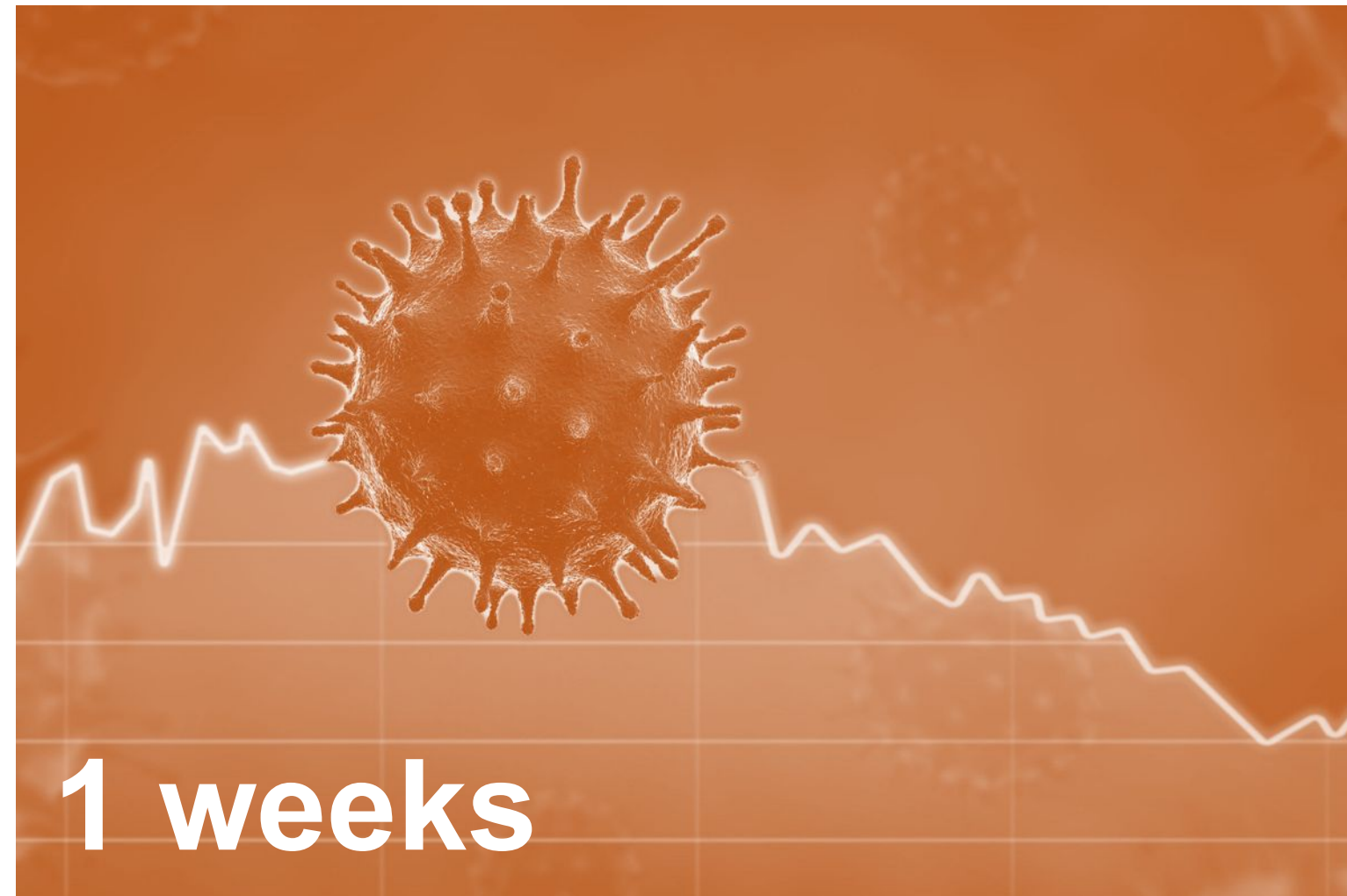
Adapting to forecasting COVID deaths



- We can't currently deal with such rapid feature drift (maybe we never can)
- So once data arrived we created a new model for the emergence of COVID



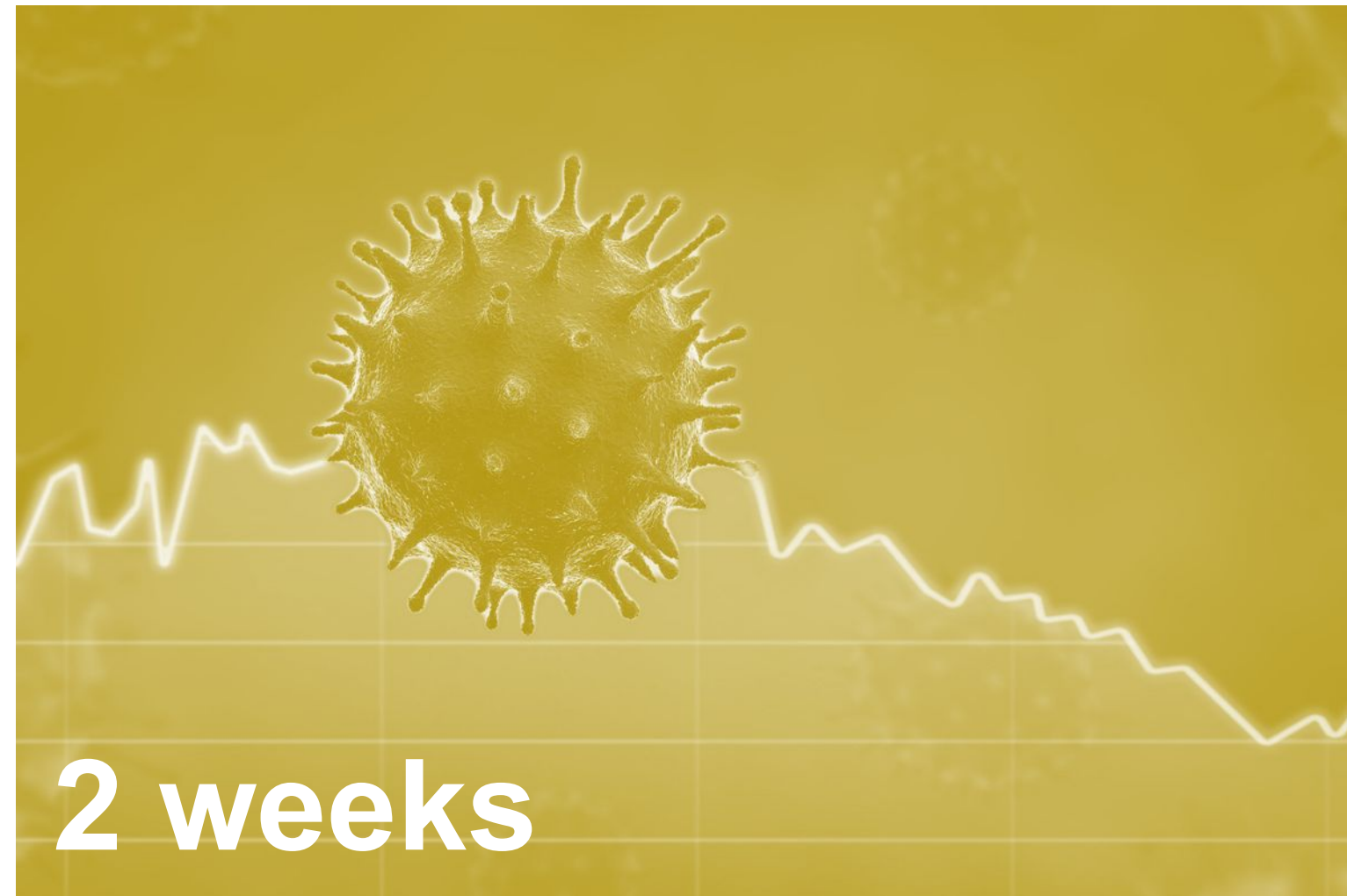
Forecasting 1 week ahead:



Forecasting deaths in a local authority 1 WEEK ahead							
Feature Set	D	C	S	DS	CS	S+	CS+
Best Model	RF	RF	XGB	XGB	XGB	RF	XGB
R ²	0.71	0.41	0.35	<u>0.73</u>	0.53	0.43	0.62
MAE	3.51	4.65	5.04	3.31	4.38	4.61	3.86

- Deaths “this week” is an excellent predictor for deaths “next week”
- Cases are a much worse indicator - the pandemic has been volatile and time-to-death > 1 week
- Sales *by themselves* are a poor indicator - for the same reasons as cases.
- BUT, as with flu prediction, sales can “fine tune” predictions of load on the NHS.

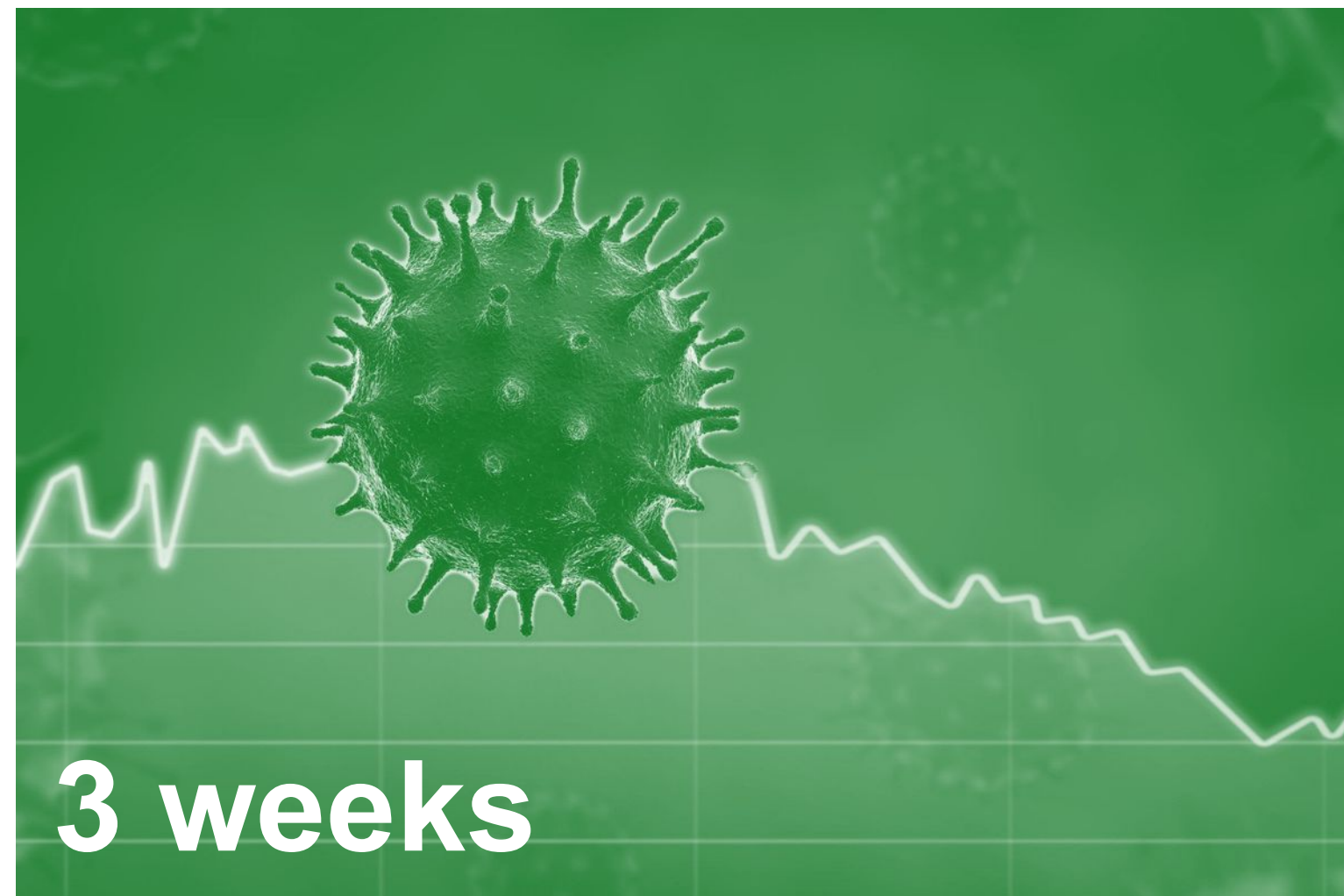
Forecasting 1 week ahead:



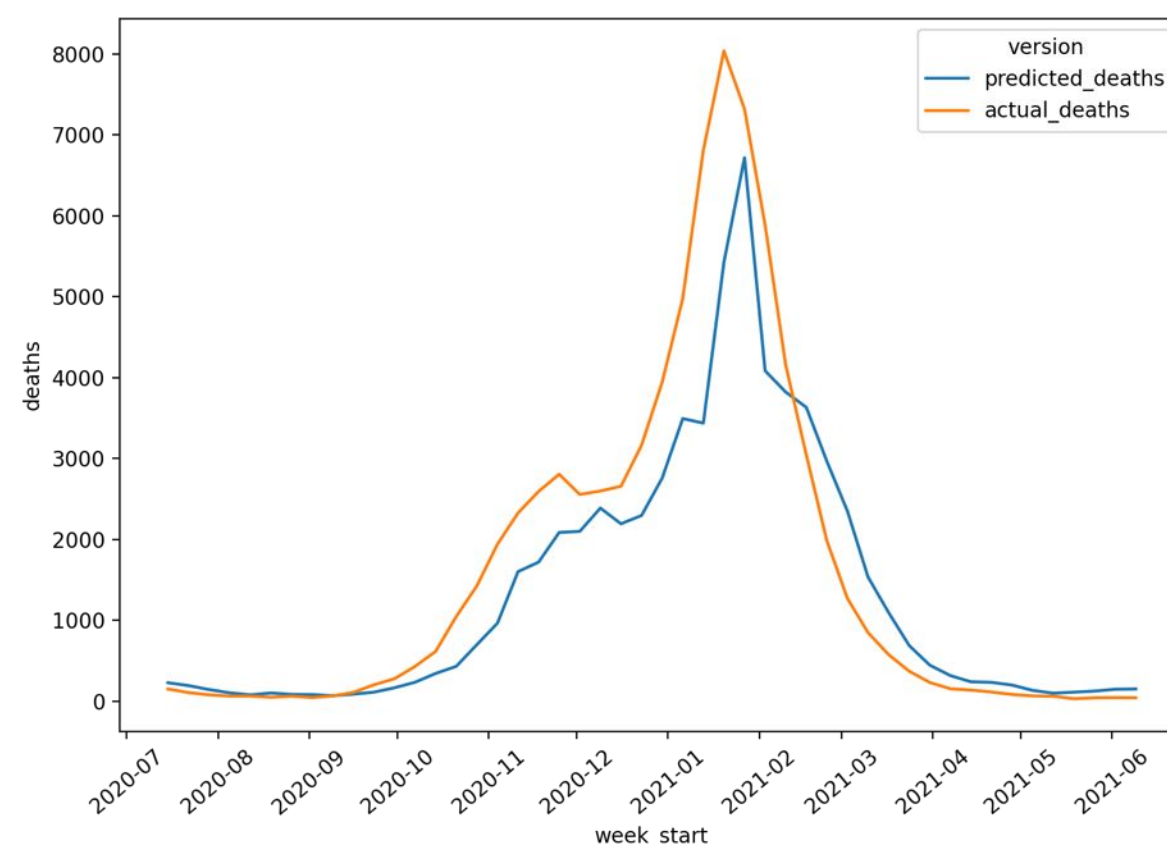
Forecasting deaths in a local authority 2 WEEKS ahead							
Feature Set	D	C	S	DS	CS	S+	CS+
Best Model	SVR	RF	RF	RF	XGB	XGB	XGB
R ²	0.47	0.50	0.42	0.60	0.62	0.48	<u>0.67</u>
MAE	4.78	4.40	4.74	4.11	3.88	4.40	3.52

- Cases are now not only a better indicator than before, and in fact **better than deaths**.
- Sales have also improved as a predictor.... but still not quite as good as cases.
- **But cases and sales TOGETHER** are now the **strongest predictor** (In particular when you add additional contextual and traditional demographic data into the analysis).

Forecasting 3 weeks ahead:



Forecasting deaths in a local authority 3 WEEKS ahead							
Feature Set	D	C	S	DS	CS	S+	CS+
Best Model	SVR	RF	RF	RF	XGB	XGB	XGB
R ²	0.26	0.44	0.44	0.52	0.61	0.50	0.68
MAE	5.75	4.91	4.75	4.52	4.11	4.41	3.67



- This pattern strengthens even further at 3 weeks ahead.
- Sales are now just as good a predictor as Cases.
- However, combined they provide even greater accuracy than 2 weeks ahead (echoing the respiratory analysis results). **3 weeks seems like a “sweet spot”**.
- **Sales clearly contains unique health information**
- capturing different demographics / types of vulnerability?



University of
Nottingham

UK | CHINA | MALAYSIA



University of
BRISTOL

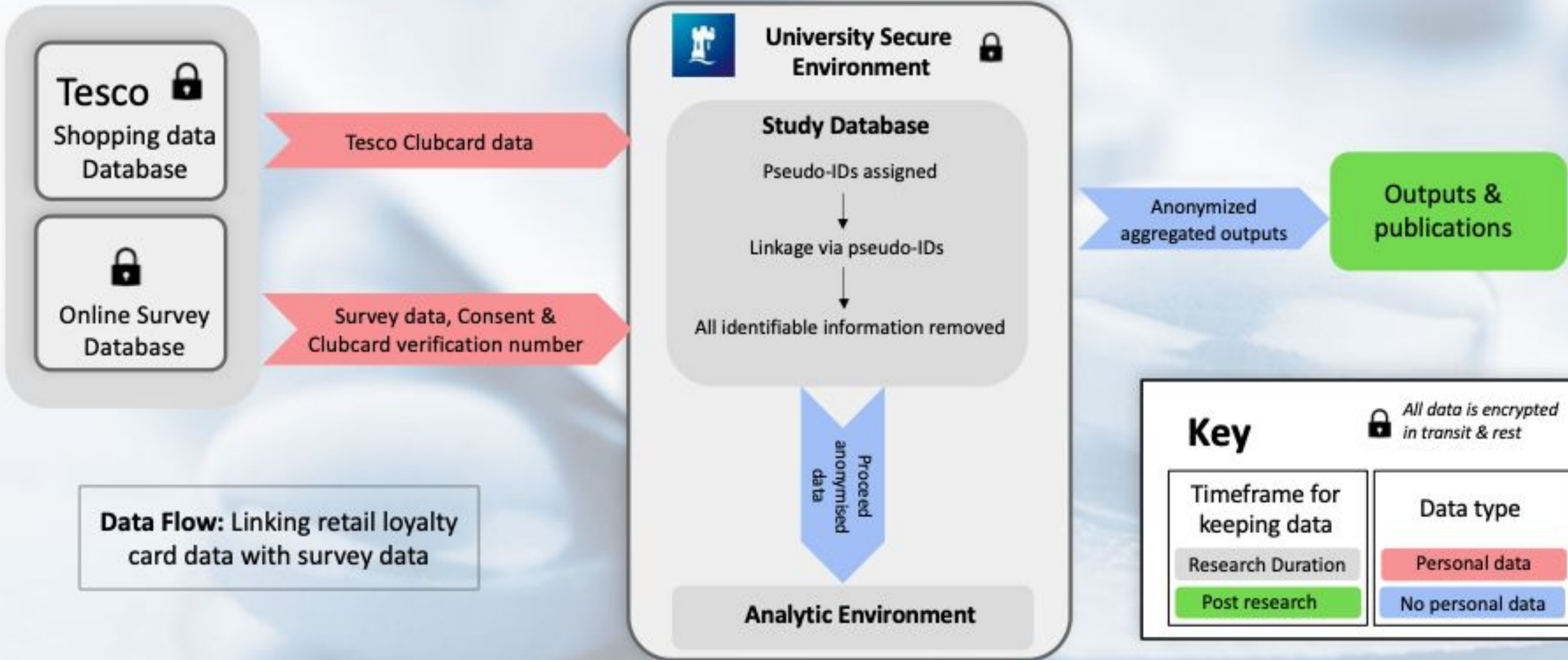
A large, high-resolution image of the Earth as seen from space, showing the Western Hemisphere. The Earth is centered in the frame, with the Atlantic Ocean, North America, and South America visible. The image has a blue and green color palette, with a white border around it.

Levelling-up via data linkages

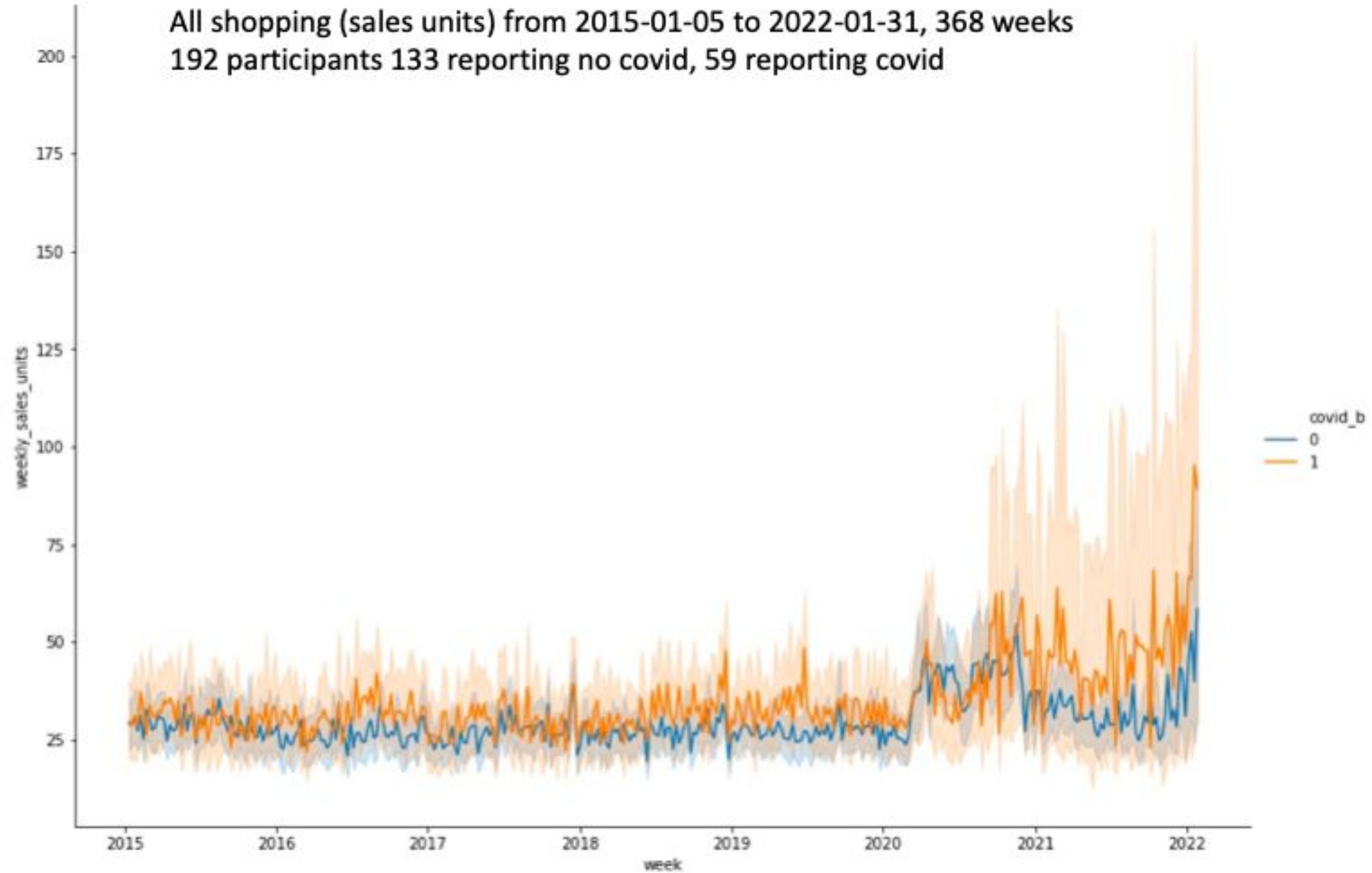
N / LAB

Digital
Footprints

Diagram summarising data flow in linking Loyalty Card data to survey data



Individual level data: COVID



Linking shopping data into cohorts

Stage 1: Data linkage

- ALSPAC
- Acceptability
- Ethical and legal basis
- Linkage infrastructure
- Data management

Stage 2: Validating data

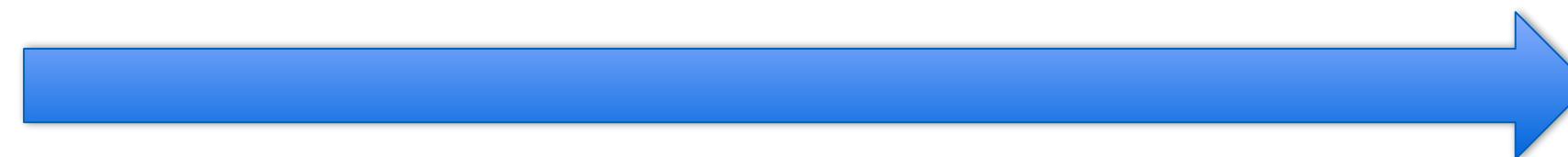
- Sampling biases
- Measurement error
- Validating patterns in the data

Stage 3: Research

- Reproductive health
- Nutrition & lifestyle
- Respiratory illness
- Self-medication



Longitudinal Cohort Studies



Other retail and transaction data



Population Health



Do you want to be involved?



DDAC

Home Our Vision Our Members Research Papers
Useful Articles Blog

DDAC is a collaboration between:

The University of Nottingham University of BRISTOL Imperial College London KING'S College LONDON

UNIVERSITY OF LEEDS WARWICK THE UNIVERSITY OF WARWICK The Alan Turing Institute UKRI

Designing the future of data donation

The Data Donation Advisory Council is an independent body of experts in data donation research, usage and best practices.

Our goal is to create trustworthy, straightforward and secure structures for the donation, transfer and use of personal digital footprint data for social good research. Connecting individuals who truly consent, with businesses and researchers who are committed to transforming health, environment and wellbeing for the benefit of all.

[Read More](#)

Inaugural meeting
26.10.22

*Special Interest
Group*

**The
Alan Turing
Institute**

**Novel Data Linkages
for Health and Wellbeing**

`james.goulding@nottingham.ac.uk`
`georgiana.nica-avram1@nottingham.ac.uk`
`nlab.org.uk`

`anya.skatova@bristol.ac.uk`
`digifootprints.co.uk`



University of
Nottingham

UK | CHINA | MALAYSIA

Thank you